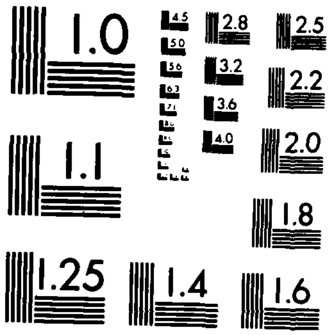
AD-A141 365 APPROPRIATENESS MEASUREMENT WITH POLYCHOTOMOUS ITEM RESPONSE MODELS ANDSS..(U) ILLINOIS UNIV AT URBANA MODEL BASED MEASUREMENT LAB F DRASGOW ET AL. APR 84 1/1

UNCLASSIFIED MEASUREMENT SER-84-1 N00014-79-C-0752 F/G 5/10 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

0

Appropriateness Measurement with Polychotomous Item Response Models and Standardized Indices

Fritz Drasgow, Michael V. Levine

and Esther A. Williams

Department of Educational Psychology
University of Illinois
Champaign, IL 61820

April, 1984

Prepared under Contract No. N00014-79C-0752, NR 154-445
and No. N00014-83K-0397, NR 150-518
Navy Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research

84 05 21 150

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
Measurement Series 84-1

2. GOVT ACCESSION NO.
AD-A141 365

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Appropriateness Measurement with Polychotomous Item Response Models and Standardized Indices

5. TYPE OF REPORT & PERIOD COVERED
Technical Report

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Fritz Drasgow, Michael V. Levine and Esther A. Williams

8. CONTRACT OR GRANT NUMBER(s)
N00014-79C-0752
N00014-83K-0397

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Model Based Measurement Laboratory
University of Illinois
Urbana, IL 61820

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
61153N RR042-04
NR 154-445
NR 150-518

11. CONTROLLING OFFICE NAME AND ADDRESS
Personnel and Training Research Programs
Office of Naval Research (code 442PT)
Arlington, VA 22217

12. REPORT DATE

13. NUMBER OF PAGES
60

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)

15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release: distribution unlimited.
Reproduction in whole or in part is permitted for any purpose of the United States Government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Latent trait theory, item response theory, multiple choice test, appropriateness measurement, person fit, polychotomous models, appropriateness index, wrong answers.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The test scores of some examinees on a multiple-choice test may not provide satisfactory measures of their abilities. The goal of appropriateness measurement is to identify such individuals. Earlier theoretical and experimental work considered examinees answering all, or almost all, test items. This article reports research that extends appropriateness measurement methods to examinees with moderately high nonresponse rates. These methods treat nonresponse as if it were a deliberate option choice and then attempt to measure the "appropriateness" of the pattern of option choices. Earlier studies used

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

only the dichotomous pattern of right and "not right" answers. A general polychotomous model is introduced along with a technique called "standardization" designed to reduce the observed confounding between measured appropriateness and ability. A standardized appropriateness index based on a polychotomous model yielded higher rates of detection of simulated spuriously low examinees than the analogous index based on a dichotomous model. However, the converse was true for simulated spuriously high examinees. Standardization was found to reduce greatly the interaction between ability and measured appropriatemess.

Appropriateness Measurement with Polychotomous Item Response Models and Standardized Indices

Fritz Drasgow, Michael V. Levine and Esther A. Williams

University of Illinois

Running head: Appropriateness Measurement

Address correspondence to: Fritz Drasgow
Department of Psychology
University of Illinois
603 E. Daniel Street
Champaign, IL 61820
USA

Abstract

The test scores of some examinees on a multiple-choice test may not provide satisfactory measures of their abilities. The goal of appropriateness measurement is to identify such individuals. Earlier theoretical and experimental work considered examinees answering all, or almost all, test items. This article reports research that extends appropriateness measurement methods to examinees with moderately high nonresponse rates. These methods treat nonresponse as if it were a deliberate option choice and then attempt to measure the "appropriateness" of the pattern of option choices. Earlier studies used only the dichotomous pattern of right and "not right" answers. A general polychotomous model is introduced along with a technique called "standardization" designed to reduce the observed confounding between measured appropriateness and ability. A standardized appropriateness index based on a polychotomous model yielded higher rates of detection of simulated spuriously low examinees than the analogous index based on a dichotomous model. However, the converse was true for simulated spuriously high examinees. Standardization was found to reduce greatly the interaction between ability and measured appropriateness.

Appropriateness Measurement with Polychotomous Item Response Models and Standardized Indices

Fritz Drasgow, Michael V. Levine and Esther A. Williams

1. Introduction

An examinee's score on a standardized multiple choice test may fail to provide a useful measure of ability for various reasons. The score may be too high because the examinee began the test with memorized answers to several questions or because the examinee copied answers to several questions from a much brighter examinee. The score may be too low because the examinee (a) made an alignment error over a block of items, answering, say, the eleventh item in the tenth space, the twelfth item in the eleventh space, . . .; (b) interpreted several very easy items in creative ways and came to well-reasoned, albeit scored-as-incorrect, answers; (c) was tested in an unfamiliar language; (d) failed to answer items on which he/she was able to eliminate several incorrect options; or, (e) worked with extreme care and consequently never reached easy items on a power test.

In all of these examples, the examinee often produces an unusual pattern of answers with relatively many easy items answered incorrectly and hard items answered correctly. Appropriateness measurement (Levine and Rubin, 1979; Levine and Drasgow, 1982,1983a; Drasgow, 1982; Hulin, Drasgow and Parsons, 1983, Chapter 4) is a model-based attempt to control test pathologies by recognizing unusual patterns. A model is fit to the item response patterns of a large sample of presumably normal examinees. Subsequently, individual examinees and their response patterns can be ordered according to how well they are fit by the group model.

Earlier appropriateness measurement work was based on models for dichotomous data and therefore was limited in two important ways. The pattern of nonresponse, which may have high diagnostic value, was ignored. In fact, earlier studies were forced to exclude examinees with high rates of nonresponse or introduce ad hoc corrections for omitting. Secondly, the earlier studies failed to take cognizance of which wrong option was chosen and therefore probably were not as sensitive to some irregularities as they might have been.

The work reported in this paper is intended to advance appropriateness measurement in two ways. A method is introduced for extending appropriateness measurement to samples of examinees with moderately high rates of nonresponse. Simultaneously methods sensitive to option choice are introduced. It will be seen that in pursuing these goals, progress has been made in comparing the appropriateness of scores at different ability levels.

2. Review of Appropriateness Measurement Terminology, Findings and Problems

The goal of appropriateness measurement is to identify examinees with inappropriate scores solely from their response patterns. This is done in two steps. First, a general psychometric model is fit to a large sample of nominally normal examinees. Then an index of goodness of fit or appropriateness index is used to measure the degree to which each individual examinee's response pattern fits the model used to characterize normal behavior.

In the first large scale, systematic appropriateness measurement study, Levine and Rubin (1979) showed that under ideal conditions certain test anomalies were detectable. They modified simulated item response data to create answer sheets with spuriously high and spuriously low scores. Three types of appropriateness indices were found to classify normal and moderately aberrant examinees rather well. However, their study was limited to simulated data conforming the "three parameter logistic model" (Birnbaum, 1968). Furthermore, their use of simulation item parameters (rather than estimated item parameters) left open the question of how well appropriateness measurement would perform in applications requiring parameter estimation.

Levine and Drasgow (1982,1983a) extended the basic results to more realistic conditions. They used actual and simlated data to study systematically the effects of the unavoidable inclusion of aberrant examinees in samples of nominally normal examinees and of errors in estimating item parameters. They found good aberrance detection with actual and simulated data despite model misspecification and parameter estimation error. However, like Levine and Rubin they only considered examinees who answered all or nearly all items and they ignored which wrong answer was chosen when a wrong answer was chosen.

The research reported in this paper extends earlier appropriateness measurement studies to (1) examinees with substantial nonresponse rates by using (2) polychotomous models and (3) standardized indices. Nonresponse to an *A*-option multiple choice item is coded as the "choice" of an *A*+1st option. In this way, every item is, in a technical sense, answered. A polychotomous psychometric model is developed to quantify the probability of each option choice (including the *A*+1st) at each ability level. Aberrance is measured by the goodness of fit of the polychotomous model.

Concern for omitting has focused our attention on conditional distributions of indices and the problems of comparing appropriateness index values at different ability levels. Examinees omitting different items are, in effect, taking different tests. A low appropriateness index value in the presence of substantial omitting may be less indicative of aberrance than a higher index value for another examinee with a different nonresponse pattern. Thus we have attempted to introduce a "common metric" for indices. Our strategy has been to divide examinees into relatively homogeneous groups by using a gross feature of the response pattern, to approximate the distribution of the index values within each group, and to use the approximated conditional distributions to define a transformation of indices to a common distribution. The "gross feature" is the maximum likelihood ability estimate, which we expected to reflect omitting rates. The common distribution was the standard normal. This process, which we call standardization, has been useful in controlling the confounding of ability and appropriateness.

3. Option Response Functions and a Constant Ability, Polychotomous Model

As a descriptive model for normal test taking behavior we have used the most general unidimensional, locally independent constant ability model that generalizes the three-parameter logistic model. It can be shown that any unidimensional model with three parameter logistic item characteristic curves and conditionally independent item responses is a special case of this model, which we call the histogram model.

To express the assumptions of the histogram model let

$$\underset{\sim}{V} = \langle \underline{V}_1, \underline{V}_2, \ldots, \underline{V}_n \rangle$$

denote the random vector of option choices and

$$\underset{\sim}{v} = \langle \underline{v}_1, \underline{v}_2, \ldots, \underline{v}_n \rangle$$

be a vector of constants indicating option choices. It is assumed that for a unidimensional ability random variable θ

$$\begin{aligned} &\text{Prob}\{\underline{V}_1=\underline{v}_1 \;\&\; \underline{V}_2=\underline{v}_2 \;\&\; \ldots \;\&\; \underline{V}_n=\underline{v}_n \mid \theta=\underline{t}\} \\ &= \prod_{i=1}^{n} \text{Prob}\{\underline{V}_i=\underline{v}_i \mid \theta=\underline{t}\} \end{aligned} \tag{1}$$

for all $\underset{\sim}{v}$ and real $\underline{t}$. Furthermore, it is assumed that if $\underline{v}_i^*$ is the correct option choice for the $\underline{i}^{\underline{th}}$ item, then for some $\underline{a}_i, \underline{b}_i, \underline{c}_i$ and for all $\underline{t}$

$$\text{Prob}\{\underline{V}_i=\underline{v}_i^* \mid \theta=\underline{t}\} = \underline{c}_i+(1-\underline{c}_i)\{1 + \exp[-\underline{a}_i(\underline{t}-\underline{b}_i)]\}^{-1} . \tag{2}$$

This model is tentatively introduced, not as a plausible model for test taking behavior, but as an admittedly crude descriptive model for test data that may or may not adequately support the extension of appropriateness measurement techniques to polychotomous data with high omitting rates. The functions

$$\underline{P}_{ij}(\underline{t}) = \text{Prob}\{\text{option } \underline{j} \text{ is chosen on item } \underline{i} | \theta=\underline{t}\}, \quad \underline{j}=1,\ldots,\underline{A}+1 \qquad (3)$$

generalize the item response function of item response theory and are cal option response functions. Their estimation is discussed in the followi section.

The likelihood of a response pattern can be easily expressed in terms of the option response functions. According to this model, the probability of sampling an examinee with response pattern $\underset{\sim}{V} = \underset{\sim}{v}$ from the subpopulation of all examinees with ability $\theta = \underline{t}$ is

$$P(\underset{\sim}{V}=\underset{\sim}{v}|\theta=\underline{t}) =$$

$$\prod_{i=1}^{n} \sum_{j=1}^{A+1} \delta_j(\underline{v}_i)\underline{P}_{ij}(\underline{t}) , \qquad (4)$$

where the first $\underline{A}$+1 positive integers are used as scores for option choices and $\delta_j(\underline{k}) = 1$ if $\underline{k} = \underline{j}$ and zero otherwise.

This equation has been used to compute polychotomous maximum likelihood ability estimates. The dichotomous model ability estimates $\hat{\theta}_d$ are obtained by maximizing the dichotomous model likelihood function

$$\prod_{i=1}^{n} [\underline{u}_i\underline{P}_i(\underline{t}) + (1-\underline{u}_i)\underline{Q}_i(\underline{t})] \qquad (5)$$

where $\underline{u}_i$ is one or zero according to whether $\underline{v}_i$ is the correct option, $\underline{P}_i(\cdot)$ is the option response function of the correct option given in equation (2), and $\underline{Q}_i(\underline{t}) = 1 - \underline{P}_i(\underline{t})$.

4. Option Curve Estimation

Various techniques have been proposed for estimating option response functions. Bock (1972) selects a parametric form for the functions and computes maximum likelihood estimates. The results of Lord (1969,1970), Samejima (1981), and Levine (1982) on ability density estimation are relevant since option characteristic curves can be represented as ratios of ability density estimates. In particular, the probability of sampling an examinee choosing the $\underline{j}^{\underline{th}}$ option from the subpopulation of examinees with ability $\underline{t}$ can be written as

$$\underline{P}_{ij}(\underline{t}) = \overline{\underline{P}}_{ij}\underline{f}_{ij}(\underline{t}) \div \underline{f}(\underline{t})$$

where $\overline{\underline{P}}_{ij}$ is the proportion of examinees choosing option $\underline{j}$ for item $\underline{i}$, $\underline{f}(\underline{t})$ is the probability density of θ and $\underline{f}_{ij}(\underline{t})$ is the θ density in the subpopulation selecting option $\underline{j}$ for item $\underline{i}$. Thus, if ability distributions can be estimated (from the dichotomously scored data), then option response curves can be estimated with no further specification of their form.

An option response function is simply the regression item option score on ability, i.e.

$$\underline{P}_{ij}(\underline{t}) = E\{\delta_j(\underline{V}_i) | \theta = \underline{t}\} \tag{6}$$

where, as before, $\delta_j(\underline{k}) = 1$ if $\underline{j} = \underline{k}$ and zero otherwise.

We have taken the simple expedient of using large sample estimates of the regression of option score on <u>estimated</u> ability $\hat{\theta}$

$$E\{\delta_j(\underline{V}_i) | \hat{\theta} = \underline{t}\}$$

as $\underline{P}_{ij}(t)$ estimates in this initial study. To obtain these estimates,

maximum likelihood ability estimates ($\hat{\theta}_d$'s) were computed for a large sample of examinees from dichotomously scored data. The examinees were grouped according to their $\hat{\theta}_d$'s . The proportion choosing each option for each ability group was used as an estimate of a point on an option response function. Linear interpolation was used between estimated points. Numerical details are given in Levine and Drasgow (1983b).

These crude estimates of option response functions are not consistent and will lead to systematic errors in ability estimates. Nonetheless, they permit us to begin an evaluation of appropriateness measurement strategies without first undertaking a major parameter estimation task.

5. The Indices and their Standardizations

In this report we are exclusively concerned with generalizations of the linear function of item scores

$$\ell_o = \sum_{i=1}^{n} \underline{u}_i \log \underline{P}_i(\hat{\theta}_d) + (1-\underline{u}_i) \log \underline{Q}_i(\hat{\theta}_d) \ . \qquad (7)$$

Here $\underline{u}_i$ is the dichotomous item score which is one if item $\underline{i}$ is answered correctly and zero if it is answered incorrectly and $\hat{\theta}_d$ maximizes the dichotomous model likelihood function. ℓ_o has the advantage of being fairly easy to compute. In comparative studies it was found to perform roughly as well as more elaborate indices (Drasgow, 1982; Levine & Rubin, 1979).

ℓ_o is the maximum of the logarithm of the dichotomous model likelihood function. The obvious generalization of ℓ_o to the histogram model is the maximum for the polychotomous model log likelihood function

$$\max_{\theta} \sum_{i=1}^{n} \sum_{j=1}^{A+1} \delta_j(\underline{v}_i) \log \underline{P}_{ij}(\theta) \ . \qquad (8)$$

In as much as our histogram model likelihood function does not have a continuous first derivative and was complicated to maximize, the generalization

$$\ell_{o,h} = \sum_{i=1}^{n} \sum_{j=1}^{A+1} \delta_j(\underline{v}_i) \log \underline{P}_{ij}(\hat{\theta}_d) \qquad (9)$$

was used. $\ell_{o,h}$ is the logarithm of the histogram model likelihood function evaluated at the dichotomous model maximum likelihood ability estimate $\hat{\theta}_d$.

As discussed in detail in Section 7 below, the distribution of ℓ_o was found to depend on ability. Therefore, two new indices were defined:

$$\underline{z}_3 = [\ell_o - E_3(\hat{\theta}_d)] \div \sigma_3(\hat{\theta}_d) \tag{10}$$

and

$$\underline{z}_h = [\ell_{o,h} - E_h(\hat{\theta}_d)] \div \sigma_h(\hat{\theta}_d) \ . \tag{11}$$

In these formulas E_3, E_h, σ_3 and σ_h are conditional means and standard deviations for the three-parameter logistic and histogram model. $E_3(\underline{t})$ is the conditional expected value of the random variable $\underline{X}_3(\underline{t})$

$$X_3(t) = \sum_{i=1}^{n} \underline{U}_i \log \underline{P}_i(\underline{t}) + (1-\underline{U}_i) \log \underline{Q}_i(\underline{t}) \tag{12}$$

computed using the three-parameter logistic model. Thus,

$$E_3(\underline{t}) = E\{\underline{X}_3(\underline{t}) | \theta=\underline{t}\} = \sum_{i=1}^{n} \underline{P}_i(\underline{t}) \log \underline{P}_i(\underline{t}) + \underline{Q}_i(\underline{t}) \log \underline{Q}_i(\underline{t}) \ . \tag{13}$$

$\sigma_3(\underline{t})$ is the square root of the conditional variance

$$\sigma_3^2(\underline{t}) = \mathrm{Var}\{\underline{X}_3(\underline{t}) | \theta=\underline{t}\} = \sum_i \underline{P}_i(\underline{t}) \underline{Q}_i(\underline{t}) [\log (\underline{P}_i(\underline{t}) / \underline{Q}_i(\underline{t}))]^2 \ . \tag{14}$$

Similarly

$$E_h(\underline{t}) = E\{\sum_{i=1}^{n} \sum_{j=1}^{A+1} \delta_j(\underline{V}_i) \log \underline{P}_{ij}(\underline{t}) | \theta=\underline{t}\} = \sum_{i=1}^{n} \sum_{j=1}^{A+1} \underline{P}_{ij}(\underline{t}) \log \underline{P}_{ij}(t) \tag{15}$$

and

$$\sigma_h^2(\underline{t}) = \mathrm{Var}\{\sum_{i=1}^{n} \sum_{j=1}^{A+1} \delta_j(\underline{V}_i) \log \underline{P}_{ij}(\underline{t}) | \theta=\underline{t}\}$$

$$= \sum_i [\sum_j \sum_k \underline{P}_{ij}(\underline{t}) \underline{P}_{ik}(\underline{t}) \log \underline{P}_{ij}(\underline{t}) \log (\underline{P}_{ij}(\underline{t}) / \underline{P}_{ik}(\underline{t}))] \ . \tag{16}$$

These transformations were found to reduce greatly the dependence of ℓ_o and $\ell_{o,h}$ on ability. Their rationale is discussed in Section 7 below.

6. Data and Parameter Estimation

Responses of approximately 75,000 examinees to the April, 1975 Scholastic Aptitude Test-Verbal section (SAT-V) were obtained from the College Entrance Examination Board. A spaced sample of 3,000 response vectors was formed by selecting the responses of every twentieth examinee, beginning with the first examinee. Item responses were then scored as correct, incorrect, omitted, or not-reached and the LOGIST computer program (Wood & Lord, 1976; Wood, Wingersky & Lord, 1976) was used to estimate item and person parameters of the three-parameter logistic model. Version 2.B of LOGIST and its default options were used. Convergence was obtained before the maximum number of iterations was reached.

These item parameter estimates were then used to construct histograms summarizing the pattern of option selection at various ability levels for the 49,470 examinees following the first 25,000 examinees in the data set. The ability estimate ($\hat{\theta}_d$) was computed for each of these examinees by the method of maximum likelihood for the dichotomously scored data. The item parameters estimated by LOGIST were used in these calculations. Omitted and not-reached items were ignored in the calculations; Lord's (1974) modified likelihood function was not used. Examinees were sorted into 25 ability strata based on estimated ability. The fourth, eighth, twelfth, . . . , 96th percentile points from the standard normal distribution were used as cutting scores to form the ability categories. The frequencies of option selection were determined for each of the 85 SAT-V items in each ability category. Finally, these frequencies were converted to proportions. The proportion choosing option $\underline{j}$ of item $\underline{i}$ for, say, the lowest ability group was taken as the estimate of $\underline{P}_{ij}(\theta)$ for $\theta = -2.054$, the second percentile of the standard normal distribution.

Proportions choosing option $\underline{j}$ of item $\underline{i}$ for the other 24 ability groups were taken as estimates of the values of $\underline{P}_{ij}$ at the 6th, 10th, . . . , 98th percentile points of the standard normal distribution. Linear interpolation between estimated values of $\underline{P}_{ij}$ was used when an ability estimate was between percentile points. No estimate of $\underline{P}_{ij}(\theta)$ was defined outside the interval [-2.054,2.054].

We chose to use 25 ability groups because this number appeared to be the best compromise between: (1) the desire to reduce sampling fluctuations by including a large number of examinees in each ability category; and (2) the desire to reduce bias by averaging over a short range of abilities. Graphs of estimated curves and further details on the estimation procedure are in Levine and Drasgow (1983b).

7. Investigation of Appropriateness Indices in Samples of Normal Examinees

7.1. Samples with Unrestricted Omitting

To examine the distributions of the various appropriateness indices, three-parameter logistic maximum likelihood estimates of ability were computed for the first 500 response vectors in the data set. A total of 464 examinees had ability estimates $\hat{\theta}_d$ in the interval $-2.054 \leq \hat{\theta} \leq +2.054$. Then ℓ_o for the three parameter logistic and histogram models, z_3 , and z_h were computed for this sample of 464 nominally normal examinees.

A scatterplot of the three-parameter logistic ℓ_o index and $\hat{\theta}$ for the first 150 examinees is presented in Figure 1. The darkened circles plotted in Figure 1 are conditional means of ℓ_o for the subset of the 464 examinees with $\hat{\theta}_d \in [t-.3, t+.3]$, $t = -2.0, -1.8, \ldots, +2.0$. The dependence of ℓ_o on estimated ability is apparent in this figure. The mean ℓ_o for examinees with $\hat{\theta}_d$ less than -1.62 is approximately -36. For examinees with $\hat{\theta}_d$ near -1.0, the mean ℓ_o is about -42. Mean ℓ_o rises as ability increases, until it reaches roughly -30 for examinees with $\hat{\theta} > +1.64$. Thus an ℓ_o score of -42 is quite low in one group of normal examinees and average in a less able group of normal examinees.

Insert Figure 1 about here

Figure 1 shows that some adjustment of the three-parameter logistic $\underline{\ell}_o$ index is necessary: the regression of $\underline{\ell}_o$ on $\hat{\theta}_d$ is not a horizontal line. Because the conditional distribution of $\underline{\ell}_o$ given $\hat{\theta}_d$ varies as a function of $\hat{\theta}$, it is difficult to interpret the magnitude of $\underline{\ell}_o$ directly.

The histogram $\ell_{o,h}$ index is plotted against $\hat{\theta}_d$ in Figure 2 for the first 150 examinees. The darkened circles are conditional means computed in the same way as in Figure 1. The dependence of histogram $\underline{\ell}_o$ is even more apparent in this figure.

Insert Figure 2 about here

If $\hat{\theta}_d$ were equal to θ, the local independence assumption could be used to reduce the dependence of ℓ_o on ability. According to the local independence assumption, in the subpopulation of examinees with ability $\theta = \underline{t}$ the item scores $\underline{u}_i$ are independent. Therefore the sum $\underline{X}_3(\underline{t})$ given in equation (12) is approximately normal with mean and variance $E_3(\underline{t})$ and $\sigma_3^2(\underline{t})$ given in formulas (13) and (14). Therefore, $\underline{z}_3$ in equation (10) would be approximately normal (0,1) for both low and high ability examinees. This was the rationale supporting the indices $\underline{z}_3$ and $\underline{z}_h$. The final expressions in equations (13) through (16) provide approximations to the actual moments when parameter estimates are substituted for parameters. Of course $\hat{\theta}_d$ is

not equal to θ , so the standardization is, at best, approximate and the above argument is merely heuristic. Nonetheless, it will be shown shortly the transformed $\underline{z}$ indices are much less sensitive to distribution of ability than the distributions of the untransformed ℓ indices.

Figure 3 presents the scatterplot of $\underline{z}_3$ and $\hat{\theta}_d$ for the first 250 examinees and conditional means obtained from the entire sample. For most values of $\hat{\theta}_d$ the conditional means are, as desired, close to the line $\underline{y} = 0$. For examinees with extreme values of $\hat{\theta}_d$, however, the conditional means are slightly less than zero.

Insert Figure 3 about here

The scatterplot of $\underline{z}_h$ and $\hat{\theta}_d$ is presented in Figure 4 for the first 250 examinees. Again, there is little relation between the standardized index and $\hat{\theta}_d$. The most striking feature of this plot is the abnormally large number of examinees with very small values of $\underline{z}_h$. In the entire sample of 464 examinees, there were 20 examinees with index scores less than -2.40; the expected number of scores in this range for a standard normal variable is only 3.8.

Insert Figure 4 about here

To determine the cause of the unusually frequent small index scores, response vectors of examinees with $\underline{z}_h$ less than -2.4 were inspected. Interestingly, many of these response vectors had very large numbers of omitted items. Of the 20 examinees with the smallest values of $\underline{z}_h$, 11 examinees (55 percent) omitted 30 or more items. In contrast, only 16 of the 444 examinees with $\underline{z}_h$ greater than -2.4 omitted 30 or more items, or 3.6 percent. Further inspection of the 27 examinees with 30 or more omits showed that their mean $\hat{\theta}_d$ was -.14 and their mean $\underline{z}_h$ was -1.87. Thus, this group appears quite ordinary with respect to ability, but has very atypical response patterns in that they omit more than 35 percent of the test.

A second group of nominally normal examinees also had very low index scores. There were seven examinees (in the sample of 464) who reached less than 77 percent of the items on the test. Their average $\underline{z}_h$ index score was -2.41.

To further investigate the relation between high omitting and $\underline{z}_h$, the response vectors of examinees 501 through 1,000 were analyzed. Note that this sample serves to replicate findings from the first sample. A total of 456 of these examinees had estimated abilities in the range from -2.05 to 2.05, 23 omitted 35 percent or more of the test items, 6 reached less than 77 percent of the test items and 16 had $\underline{z}_h$ values of less than -2.4.

The relations between omitting and $\underline{z}_h$ that were noted for the first sample of examinees were confirmed in this second sample. In particular the mean $\underline{z}_h$ for examinees who omitted more than 35 percent of the test was -1.85. Seven of these examinees had $\underline{z}_h$ values of less than -2.4. Even stronger results occurred for the six examinees who reached less than 77 percent of the test items: all six had $\underline{z}_h$ scores of less than -2.4. The mean $\underline{z}_h$ for this group was -3.08.

A total of 430 examinees in the second SAT-V sample omitted less than 35 percent of the test <u>and</u> reached 77 percent or more of the test items. In this group, 6 examinees had $\underline{z}_h$ scores of less than -2.4. The expected number for a standard normal population is 3.5. In contrast, of the 26 examinees who omitted more than 35 percent of the test, <u>or</u> reached less than 77 percent of the test items, 10 had $\underline{z}_h$ values of less than -2.4; the expected number is .2.

It is not surprising that high omitting rates and not finishing the exam cause $\underline{z}_h$ to indicate aberrance. Perhaps the most important point to note about the relation between high omitting and $\underline{z}_h$ is that it is <u>not</u> high omitting per se that causes very extreme $\underline{z}_h$ values. Instead, it is the too frequent omitting of easy items, or items with very effective

distractors. For example, examinees who reached less than 77 percent of the test items did not attempt several easy items; across the last 10 items on each of the two SAT-V subsections, there were 5 items with $\hat{b}_i$ values less than -1.0, and 9 items with $\hat{b}_i$ values less than 0.0.

Because examinees who omit more than 35 percent of the test or reach less than 77 percent of test items appear to receive spuriously low test scores, they are excluded from all subsequent analyses.

7.2. Restricted Omitting Sample

To investigate further the distributions of z_h and z_3, a large sample of nominally normal examinees was formed. First, three parameter logistic maximum likelihood estimates of ability were computed for examinees 10,001 to 14,000 on the SAT-V tape. Then, examinees who met the following three criteria were included in the sample:

1. Less than 35 percent of the test items were omitted;
2. 77 percent or more of the test items were reached;
3. Estimated ability was in the range $-2.05 \leq \hat{\theta} \leq 2.05$.

The z_h and z_3 appropriateness indices were computed for the 3478 examinees who satisfied these criteria.

Figure 5 presents the cumulative frequency distributions for z_h and z_3 in the sample of 3478 nominally normal examinees. The cumulative distribution function of the standard normal distribution is also presented. From Figure 5, it is apparent that the distributions of z_h and z_3 are slightly asymmetric: there are relatively few examinees with index scores between -2.0 and 0.0, and relatively many with index scores between 0.0 and 1.2. Both empirical distributions are significantly different form the standard normal distribution ($\alpha = .01$) by the Kolmogorov-Smirnov test.

Insert Figure 5 about here

For the purposes of appropriateness measurement, it is not essential that $\underline{z}_h$ and $\underline{z}_3$ follow standard normal distributions. It *is* important that each index be distributed similarly across values of $\hat{\theta}_d$. Table 1 presents information concerning the left tails of the distributions of $\underline{z}_3$ and $\underline{z}_h$ within five mutually exclusive ability intervals. The left tails of the conditional distributions of $\underline{z}_3$ are relatively similar across the five ability intervals. The largest difference between cumulative proportions at any cutting score is only .03. The left tails of the conditional cumulative proportions of $\underline{z}_h$ exhibit less invariance; here the largest difference is .054.

Insert Table 1 about here

The relatively large differences in conditional distributions of $\underline{z}_h$ for different ability levels may result from the presence of truly aberrant response patterns in the sample of nominally normal response vectors rather than inaccuracies in the standardization approximation. (It will be shown in Section 8 that $\underline{z}_3$ and $\underline{z}_h$ are more sensitive to some types of aberrance for examinees of very high or very low ability.) To investigate this possibility, the research described in the present subsection was replicated using simulation data.

7.3. Samples Simulated According to the Three Parameter Logistic Model and Histogram Model

Samples of 4,000 simulated examinees were generated using the three parameter logistic model and histogram model. Hypothetical probabilities of correct responses on dichotomously scored items and hypothetical probabilities of option selection on polychotomously scored items were computed using the three parameter logistic model ICC estimates and histogram option response function estimates described in Section 6. For each simulated examinee, an ability was sampled from the standard normal distribution truncated to the interval $[-2.05, 2.05]$. Responses to 85 items were then simulated as 85 independent multinomials with response probabilities obtained by substituting sampled ability in the three parameter logistic ICCs and histogram option response functions.

Ability was estimated for each response vector by the methods described in Sections 7.1 and 7.2. Response vectors for which $|\hat{\theta}_d| > 2.05$ were discarded so that the results described in this section would be comparable to the results presented in Section 7.2. The $\underline{z}_3$ and $\underline{z}_h$ indices were computed using estimated ability. Simulated examinees were then sorted into five ability categories on the basis of θ (not $\hat{\theta}_d$).

The cumulative proportions of appropriateness index scores for the five ability intervals are shown in Table 2. Note that (1) there was no model misspecification or parameter estimation problem here because the item parameters and option response probabilities used to compute index values were identical to those used to generate response vectors; and (2) there were no truly aberrant response vectors present. Although the cumulative proportions in Table 2 tend to be somewhat smaller than the corresponding proportions in

Table 1, the overall pattern is similar. Again, the largest difference in conditional proportions for the three parameter logistic is .03. The largest difference for the $\underline{z}_h$ proportions is .046, which again suggests that there is less invariance of the conditional $\underline{z}_h$ distribution than for the $\underline{z}_3$ distribution.

Insert Table 2 and Table 3 about here

Shown in Table 3 are the conditional proportions of index values obtained when simulated examinees are sorted into ability categories on the basis of $\hat{\theta}_d$ rather than θ. The cumulative proportions for $\underline{z}_3$ are similar to the proportions shown in Table 2. Curiously, $\underline{z}_h$ shows more invariance across ability categories in Table 3 than in Table 2.

7.4. Summary

The standardized ℓ_o indices, $\underline{z}_h$ and $\underline{z}_3$, have empirical distributions that are reasonably close to the standard normal distribution. The Kolmogorov-Smirnov tests indicate that $\underline{z}_h$ and $\underline{z}_3$ do not exactly follow the standard normal distribution. Furthermore, Tables 1 and 2 indicate that the distributions of $\underline{z}_h$ and $\underline{z}_3$ are not completely independent of estimated ability. However, Figure 5 and Tables 1, 2, and 3 suggest that these effects are fairly small. In addition, these tables show that high rates of detection of aberrant response vectors will not result solely from differences in ability distributions. For this reason, $\underline{z}_h$ and $\underline{z}_3$ are used as appropriateness indices in the next section.

8. Appropriateness Measurement with Standardized ℓ_0 Indices

8.1. Overview

In this section, we compare the distributions of the two appropriateness indices in samples of normal examinees to the distributions in samples of examinees whose response vectors have been modified to simulate spuriously high and spuriously low examinees. The power of an appropriateness index is indicated by the extent to which the index separates the normal and aberrant groups.

8.2. Normal and Aberrant Groups

The normal group consists of the 3,478 nominally normal, low omitting examinees with $-2.05 \leq \hat{\theta} \leq 2.05$ previously described.

The aberrant groups were formed by the following process. First, only examinees with less than 35 percent of the test omitted and 77 percent or more test items reached were considered. Then, starting with examinee 1,001 on the SAT-V tape, 300 examinees with estimated ability in each of the five ability categories were selected from the next 2,000 records. The $\hat{\theta}_d$ categories, termed quintiles, are:

Quintile	$\hat{\theta}$ range
Q1	[-2.05, -.80];
Q2	(-.80, -.24];
Q3	(-.24, .24];
Q4	(.24, .80];
Q5	(.80, 2.05].

These quintiles of response vectors were then subjected to various types of tampering to simulate aberrance.

The $k\%$ spuriously high modification consisted of randomly selecting $k\%$ of the examinee's original responses without replacement. Then each response was rescored as correct, regardless of the original response. Note that omits were treated as any other response category and rescored as correct if selected. Ten, 20, and 30% modifications were applied to each of the five quintiles.

The $k\%$ spuriously low modification was slightly more complex. First, each examinee's response vector was inspected to determine the proportion, q, of omitted items. Then $k\%$ of the examinee's original responses were selected randomly and without replacement. Each item was rescored as an omit with probability q. Options A through E were selected with probability $(1-q)/5$. Note that this procedure reflects the examinee's original propensity to omit items. Again, 10, 20, and 30% modifications were applied to the quintiles.

After tampering with the response vectors in a quintile, ability was estimated for each modified response vector using the three parameter logistic model. Then z_h and z_3 were computed for the modified response vectors in the quintile.

8.3. ROC Curves

We used ROC curves to display the effectiveness of an index for detecting simulated aberrance. Here, a value of the appropriateness index, say t, is specified. Then the proportion of normal and aberrant response vectors with index values less than t are determined. Let

$$x(t) = \text{proportion of normal examinees with index values } \leq t$$
$$y(t) = \text{proportion of aberrant examinees with index values } \leq t .$$

Plotting the $< x(t), y(t) >$ pairs for several values of t produces an

ROC curve. An ROC curve that indicates good detection of aberrance is one that rises sharply from the origin to the upper left hand corner of the plot. A random classification system would produce an ROC curve that lies along the 45 degree diagonal line. To conserve space, we only plot ROC curves for low false alarm rates: $x(t) \leq .20$. An elementary description of the use of ROC curves in appropriateness measurement is given by Hulin, Drasgow, and Parsons (1983).

8.4. Results for the Spuriously Low Modification

Figure 6 presents the ROC curves for the spuriously low modification, with panel (A) corresponding to the $\underline{z}_h$ index and panel (B) corresponding to $\underline{z}_3$ index. The 30% modification is indicated by circles, the 20% modification by squares, and the 10% modification by a solid line. The 45 degree diagonal line is also plotted.

Insert Figure 6 about here

The panels in Figure 6 portray an orderly, coherent pattern of detectability. In each case, tampering with more items leads to greater detectability. This is indicated by the ROC curves for the 30% modifications always rising more sharply than the other two curves, and the 20% modification ROC curves rising more sharply than the 10% curves.

It is clear that detectability increases with increasing ability. For example, the lowest detection rates occur for the first quintile where examinees had estimated ability in the range -2.05 to -.80 prior to tampering. It is obvious that the spuriously low treatment would have relatively little

effect on the responses of these low ability examinees. In contrast, examinees in quintile 5 all had estimated ability in the range .80 to 2.05 prior to tampering. Here the effects of the spuriously low modification on each examinee's response vector are much larger, and this is reflected in very high detection rates. Note that detectability increases evenly as pre-tampering ability increases; there is not a particular ability level below which appropriateness measurement is completely ineffective and above which appropriateness measurement is quite effective.

Despite the crude estimates of the histogram model's option response functions, it is clear that the $\underline{z}_h$ index is substantially superior to the $\underline{z}_3$ index. ROC curves for $\underline{z}_h$ generally rise more sharply than the corresponding $\underline{z}_3$ ROC curves, and hence provide better aberrance detection.

A reason that the histogram model affords better aberrance detection is straightforward: aberrant responses, as conceptualized and simulated here, are essentially random. Thus easy items can be missed and some extremely improbable (P_{ij} less than .01) incorrect options are selected. The dichotomous test model is sensitive to incorrect responses to easy items, but is insensitive to the pattern of incorrect option selection. In contrast, the $\underline{z}_h$ index is affected by the selection of an incorrect option.

8.5. Results for the Spuriously High Modification

Figure 7 presents the results for the spuriously high modification. Again, the 30, 20, and 10% modifications are indicated by circles, squares, and solid lines, respectively. Clearly, tampering with more items increases detectability. As expected, detectability decreases with increasing ability in Figure 7. Providing the answer key for, say, 20% of the exam to bright examinees has a relatively small impact on their answers, and is not likely to be detectable by appropriateness measurement techniques. In contrast, providing low

ability examinees with answers to 20% of the exam will have substantial effects on their responses. As seen in Figure 7, this type of spuriously high score is detectable, especially by the $\underline{z}_3$ index.

Insert Figure 7 about here

Perhaps the most interesting result obtained from the spuriously high modification is the finding that the $\underline{z}_3$ index is clearly superior to the $\underline{z}_h$ index. This result appears counter-intuitive because the histogram model, which provides a fuller description of the test-taking behavior of normal examinees, should provide more power in detecting departures from normal test-taking behavior. We believe that the superiority of the dichotomous test model for detecting spuriously high examinees is chiefly a result of the particular class of appropriateness indices under consideration (i.e., the $\underline{\ell}_o$ class). This class of indices is subject to a "swamping" effect when utilized to detect spuriously high response vectors on polychotomously scored multiple choice tests. Other, more sophisticated indices may not be affected similarly.

The swamping effect is perhaps best described by example. Table 4 presents the frequency distribution of the terms that compose ℓ_o and $\ell_{o,h}$ for the first examinee in Quintile 2:

$$\ell_{o,3}{}^{(i)} = \underline{u}_i \log \underline{P}_i(\hat{\theta}_d) + (1-\underline{u}_i) \log \underline{Q}_i(\hat{\theta}_d)$$

and

$$\ell_{o,h}{}^{(i)} = \sum_{j=1}^{A+1} \delta_j(\underline{v}_i) \log \underline{P}_{ij}(\hat{\theta}_d)$$

respectively.

Note that prior to tampering there was only one term less than -2.0 for the three parameter logistic model but there were 17 such terms for the histogram model. After tampering, there were 5 terms less than -2.0 for the dichotomous model (an increase of 400%) and 22 terms for the histogram model (an increase of 29%). It is interesting to note that three of the smallest four $\ell_0^{(i)}$ terms for the logistic model after tampering were items rescored as correct during tampering. In contrast, none of the three smallest histogram terms had been subjected to tampering, and only two of the smallest 11 terms had been affected by tampering.

Insert Table 4 about here

This example illustrates the swamping effect. A normal number of relatively rare incorrect option selections and mistakes on easy items—as noted above, 9 for the examinee in Table 4—camouflages correct answers to hard items produced by the spuriously high modification. This occurs because the probabilities for some incorrect options are very nearly zero in the histogram model, and most examinees choose a few of these improbable incorrect options during the 85 item SAT-V exam. Swamping occurs much less in the three parameter logistic model because the model does not differentiate between the various incorrect options.

9. Discussion

From the research presented in this article it is apparent that standardization substantially reduces the confounding between measured appropriateness and ability: The conditional distributions of $\underline{z}_3$ and $\underline{z}_h$ are more nearly invariant across ability levels than are the conditional distributions of ℓ_o and $\ell_{o,h}$. Thus, standardized index scores for examinees of different abilities can be compared more easily when making classification decisions.

It must be emphasized that our implementation of the standardization concept (i.e. transforming index values to make the conditional distribution of an appropriateness index independent of estimated ability) can be improved in many ways. An improved estimate of the conditional distribution of an index can be obtained, if not analytically then by simulation. A more "robust" estimate of ability can be obtained by reducing the relative contribution of improbable responses to the estimate (Wainer & Wright, 1980; Jones, 1982). Our studies show that standardization is needed and is easily implemented.

In pilot studies for the research described by Levine and Rubin (1979), it was found that ℓ_o provided very low rates of detection of aberrant response patterns in samples of examinees with unrestricted omitting. Levine and Rubin found much higher detection rates when samples were restricted to low rates of omitting. To handle higher rates of omitting we have introduced polychotomous models. It is interesting to note, however, that standardization of the dichotomous model appropriateness index ℓ_o allows high detection rates in samples with only weak restrictions on omitting rates. These detection rates seem to be nearly as high as detection rates for ℓ_o in samples with low omitting rates.

Improved detection of response patterns modified by the spuriously low treatment was also obtained from use of the $\underline{z}_h$ index. This index is sensitive to the pattern of incorrect option selection and consequently facilitates identification of examinees who choose unusual options when they respond incorrectly.

The $\underline{z}_h$ index was not as effective as the $\underline{z}_3$ index in identifying spuriously high response patterns. Thus, we are left with the question of whether our particular choice of a polychotomous model appropriateness index was unfortunate: Would a different polychotomous model appropriateness index provide much higher detection rates? In our current research, we have identified an optimal appropriateness index for spuriously high response patterns. Our preliminary results indicate that $\underline{z}_3$ detects spuriously high patterns at a rate much closer to the optimal index than $\underline{z}_h$, but not so well as to discourage refinements of $\underline{z}_h$ and the formulation of an alternative polychotomous index.

Omitted items are ignored when computing the standardized dichotomous appropriateness index. The standardized polychotomous index, in contrast, treats nonresponse as the selection of the ($\underline{A}$+1)th option on an $\underline{A}$ option multiple choice item. This "option" is then treated in a fashion similar to the other options when computing $\underline{z}_h$. Examinees who omitted a large number of items or who failed to reach many items frequently received very low $\underline{z}_h$ scores. Because it seems likely that these examinees would receive higher test scores (which are a linear function of number correct minus one-fourth of number incorrect) if they answered more items, it appears that $\underline{z}_h$ has identified one form of naturally occuring spuriously low test score.

The very low appropriateness scores observed among nominally normal examinees with high nonresponse rates can be seen as casting doubt on the unidimensional, local independence assumptions of the histogram model. It seems likely that there are substantial individual differences between exam-

inees in rates of responding and willingness to guess or use partial information. These departures from unidimensionality, though obvious in retrospect, constitute an serendipitous finding of considerable practical importance. Excessively conservative examinees who are reluctant to use partial information, examinees who perseverate on difficult items and other able, low scoring examinees with high nonresponse rates indeed do have inappropriately low number right scores. It seems desirable to identify and counsel them. From the testing organization's point of view, it seems wise to exclude them from item parameter estimation samples since their presence may introduce additional sampling error (and possibly bias) in item parameter estimates.

References

Birnbaum, A. (1968) Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick, Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley.

Bock, R.D. (1972) Estimating item parameter and latent ability when responses are scored in two or more nominal categories. Psychometrika, 37, 29-51.

Drasgow, F. (1982) Choice of test model for appropriateness measurement. Applied Psychological Measurement, 6, 297-308.

Hulin, C.L., Drasgow, F., & Parsons, C.K. (1983) Item response theory: Application to psychological measurement. Homewood, Ill.: Dow Jones-Irwin.

Jones, D.H. (1982) Tools of robustness for item response theory. Research Report 82-41. Princeton, N.J.: Educational Testing Service.

Levine, M.V. (1983) The trait in latent trait theory. In D.J. Weiss (Ed.), Proceedings of the 1982 Item Response Theory/Computerized Adaptive Testing Conference. Minneapolis: University of Minnesota, Department of Psychology, Computerized Adaptive Testing Laboratory.

Levine, M.V. & Drasgow, F. (1982) Appropriateness measurement: Review, critique and validating studies. British Journal of Mathematical and Statistical Psychology, 35, 42-56.

Levine, M.V. & Drasgow, F. (1983) Appropriateness measurement: Validating studies and variable ability models. In D.J. Weiss (Ed.) New horizons in testing: Latent trait test theory and computerized adaptive testing. New York: Academic Press, in press. (a)

Levine, M.V. & Drasgow, F. (1983) The relation between incorrect option choice and estimated ability. Educational and Psychological Measurement, in press. (b)

Levine, M.V. & Rubin, D.F. (1979) Measuring the appropriateness of multiple choice test scores. Journal of Educational Statistics, 4, 269-290.

Lord, F.M. (1969) Estimating true-score distributions in psychological testing (An empirical Bayes estimation problem). Psychometrika, 34, 259-299.

Lord, F.M. (1970) Item characteristic curves as estimated without knowledge of their mathematical form - A confrontation of Birnbaum's logistic model. Psychometrika, 35, 43-50.

Lord, F.M. (1974) Estimation of latent ability and item parameters when there are omitted responses. Psychometrika, 39, 247-264.

Samejima, F. (1981) Final report: Efficient methods of estimating the operating characteristics of item response categories and challenge to a new model for the multiple-choice item. Technical Report. Knoxville, Tenn.: Department of Psychology, University of Tennessee.

Wainer, H. & Wright, B.D. (1980) Robust estimation of ability in the Rasch model. Psychometrika, 45, 373-391.

Wood, R.L. & Lord, F.M. (1976) A user's guide to LOGIST. Research Memorandum 76-4. Princeton, N.J.: Educational Testing Service.

Wood, R.L., Wingersky, M.S., & Lord, F.M. (1976) LOGIST - A computer program for estimating examinee ability and item characteristic curve parameters. Research Memorandum 76-6. Princeton, N.J.: Educational Testing Service.

Acknowledgements

This work was supported by ONR contracts N00014-79C-0752, NR 154 445 and N00014-83K-0397, NR 150 518. We are grateful to the College Entrance Examination Board for providing access to the Scholastic Aptitude Test data.

Table 1

Cumulative Proportions of Appropriateness Index

Scores at Various Cutting Scores for 3478 SAT-V Examinees

Index	Cutting Score	Normal Curve	Ability Interval*				
			Low	Mod. Low	Ave.	Mod. High	High
z_3	-2.58	.005	.014	.006	.012	.007	.018
	-1.96	.025	.045	.025	.030	.028	.048
	-1.64	.050	.069	.047	.050	.055	.072
	-1.30	.097	.106	.079	.089	.085	.109
z_h	-2.58	.005	.006	.002	.002	.006	.018
	-1.96	.025	.035	.016	.008	.013	.037
	-1.64	.050	.055	.017	.020	.024	.056
	-1.30	.097	.094	.050	.040	.046	.087
Total N in Ability Interval			650	643	643	672	870

* The ability intervals are: low = [-2.05, -0.80], moderately low = (-0.80, -0.24], average = (-0.24, 0.24], moderately high = (0.24, 0.80], and high = (0.80, 2.05].

Table 2

Cumulative Proportions of Appropriateness Index

Scores at Various Cutting Scores for Simulated Examinees

Index	Cutting Score	Normal Curve	Ability Interval*				
			Low	Mod. Low	Ave.	Mod. High	High
z_3	-2.58	.005	.007	.002	.005	.003	.003
	-1.96	.025	.032	.018	.013	.015	.018
	-1.64	.050	.047	.045	.030	.035	.033
	-1.30	.097	.083	.088	.058	.081	.063
Total N in Ability Interval			748	815	762	780	795
z_h	-2.58	.005	.004	.003	.000	.001	.000
	-1.96	.025	.013	.012	.005	.008	.004
	-1.64	.050	.042	.027	.016	.022	.024
	-1.30	.097	.079	.071	.033	.054	.064
Total N in Ability Interval			745	816	762	780	791

Note: Simulated examinees were sorted into ability intervals on the basis of θ .

*The ability intervals are: low = [-2.05, -0.80], moderately low = (-0.80, -0.24], average = (-0.24, 0.24], moderately high = (0.24, 0.80], and high = (0.80, 2.05].

Table 3

Cumulative Proportions of Appropriateness Index Scores at Various Cutting Scores for Simulated Examinees

Index	Cutting Score	Normal Curve	Ability Interval*				
			Low	Mod. Low	Ave.	Mod. High	High
z_3	-2.58	.005	.005	.005	.005	.003	.001
	-1.96	.025	.024	.022	.021	.014	.017
	-1.64	.050	.038	.046	.042	.032	.032
	-1.30	.097	.062	.097	.080	.073	.061
Total N in Ability Interval			742	794	765	752	847
z_h	-2.58	.005	.004	.000	.003	.001	.000
	-1.96	.025	.011	.008	.011	.008	.005
	-1.64	.050	.034	.020	.028	.020	.027
	-1.30	.097	.065	.049	.064	.050	.073
Total N in Ability Interval			759	754	793	747	841

Note: Simulated examinees were sorted in to ability intervals on the basis of $\hat{\theta}$.

* The ability intervals are: low = [-2.05, -0.80], moderately low = (-0.80, -0.24], average = (-0.24, 0.24], moderately high = (0.24, 0.80], and high = (0.80, 2.05].

Table 4

Frequency Distribution of $\ell_o^{(i)}$ Terms

for the First Examinee in Quintile 2

Interval	Three Parameter Logistic Model	Histogram Model
Distribution for Original Responses		
(-1.0, 0.0]	74	37
(-2.0, -1.0]	9	31
(-3.0, -2.0]	1	15
(-4.0, -3.0]	0	2
Omit	1	-
	$\underline{z}_3 = 1.22$	$\underline{z}_h = 0.57$
Distribution After 20% Spuriously High Modification		
(-1.0, 0.0]	70	39
(-2.0, -1.0]	10	24
(-3.0, -2.0]	5	19
(-4.0, -3.0]	0	3
Omit	0	-
	$\underline{z}_3 = -1.96$	$\underline{z}_h = -1.09$

Figure Captions

1. Three parameter logistic ℓ_o plotted against $\hat{\theta}_d$ for 150 nominally normal examinees.

2. Histogram $\ell_{o,h}$ plotted against $\hat{\theta}_d$ for 150 nominally normal examinees.

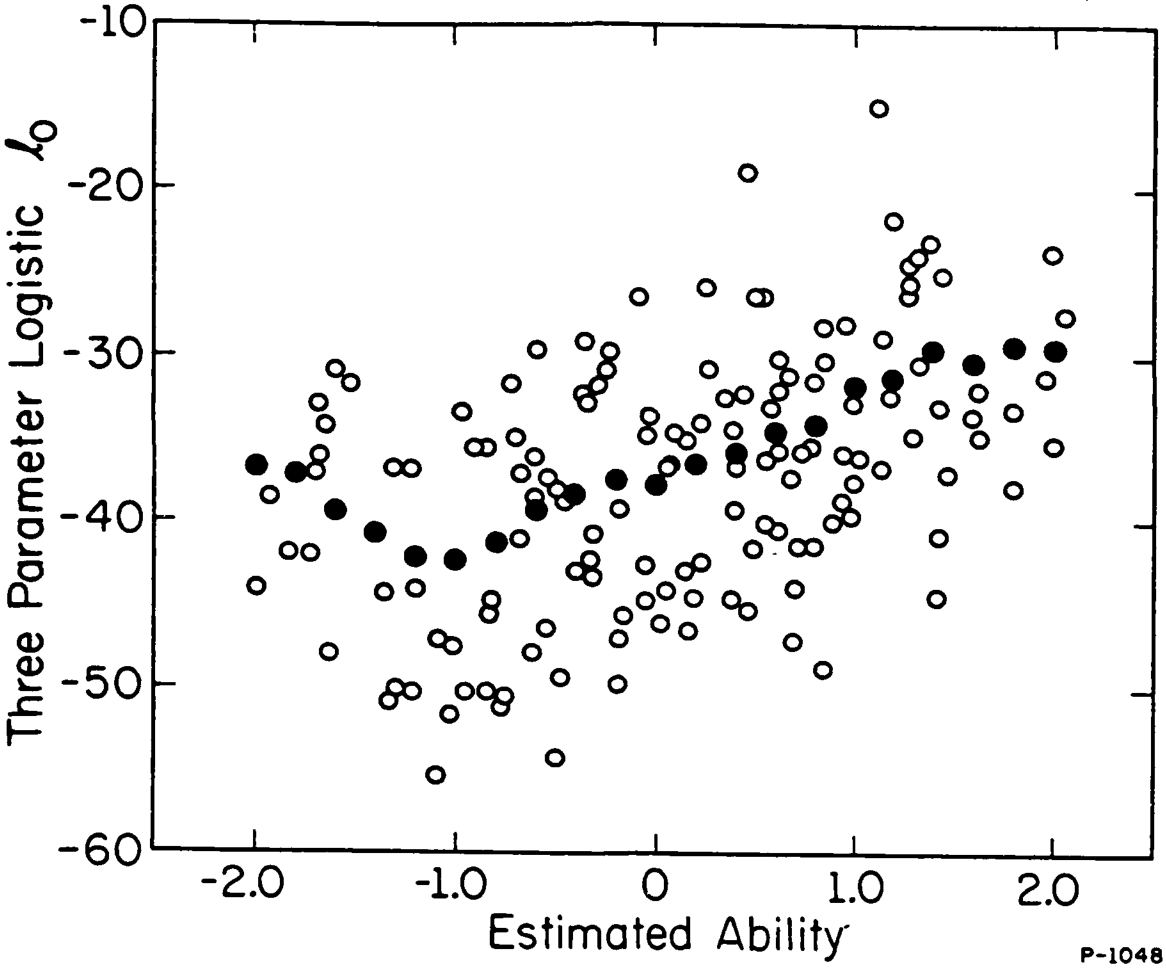
3. Standardized three parameter logistic appropriateness index $\underline{z}_3$ plotted against $\hat{\theta}_d$ for 250 nominally normal examinees.

4. Standardized histogram appropriateness index $\underline{z}_h$ plotted against $\hat{\theta}_d$ for 250 nominally normal examinees.

5. Cumulative proportions for the standard normal distribution and the standardized $\underline{z}_3$ and $\underline{z}_h$ appropriateness indices.

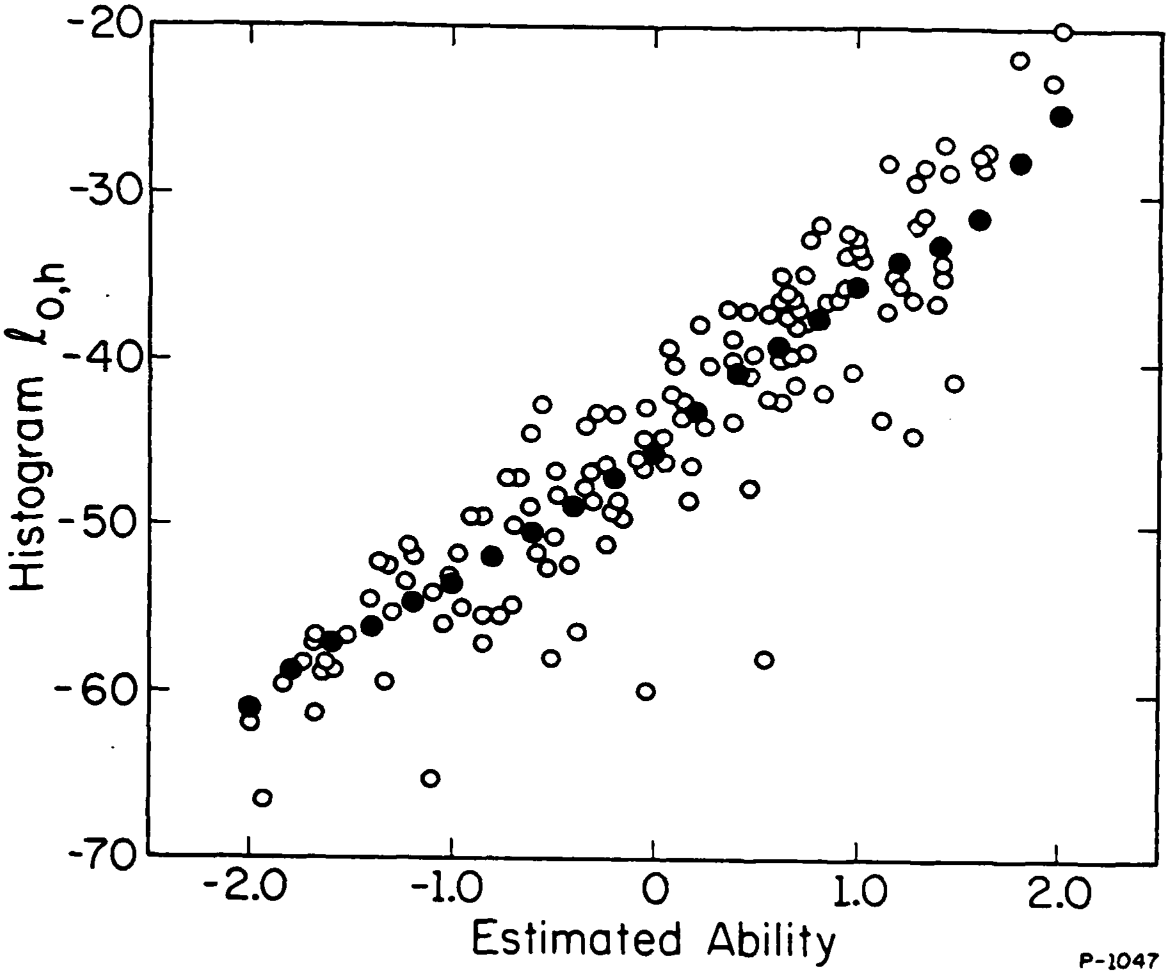
6. ROC curves for the spuriously low manipulations. Panel (A) presents results for the $\underline{z}_3$ appropriateness index and panel (B) presents results for the $\underline{z}_h$ appropriateness index.

7. ROC curves for the spuriously high manipulations. Panel (A) presents results for the $\underline{z}_3$ appropriateness index and panel (B) presents results for the $\underline{z}_h$ appropriateness index.


-10
-20
-30
-40
-50
-60
Three Parameter Logistic l_0
-2.0
-1.0
0
1.0
2.0
Estimated Ability

P-1048

Fig. 1


-20
-30
-40
-50
-60
-70
Histogram $\ell_{0,h}$
-2.0
-1.0
0
1.0
2.0
Estimated Ability

P-1047

Fig. 2

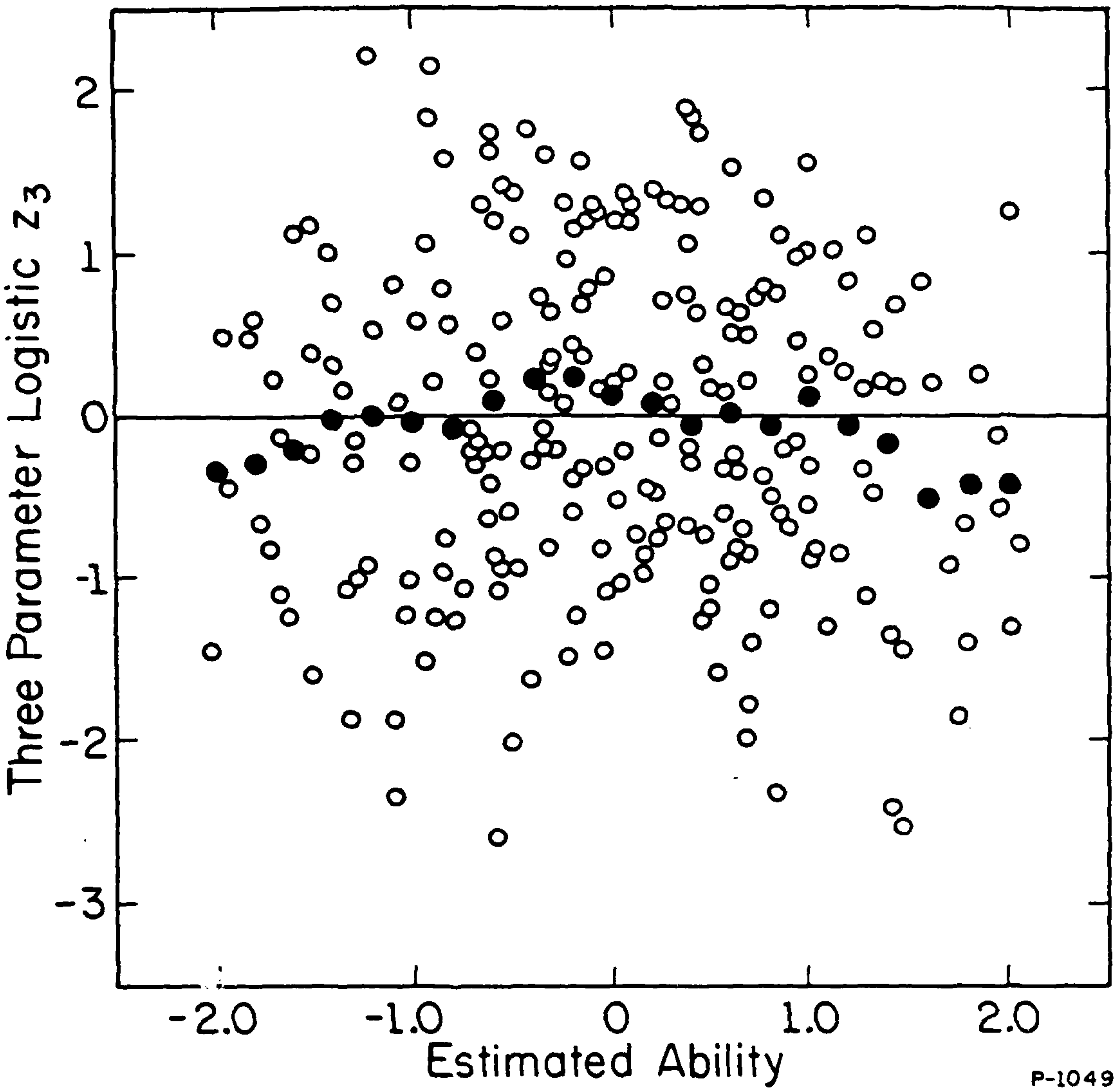
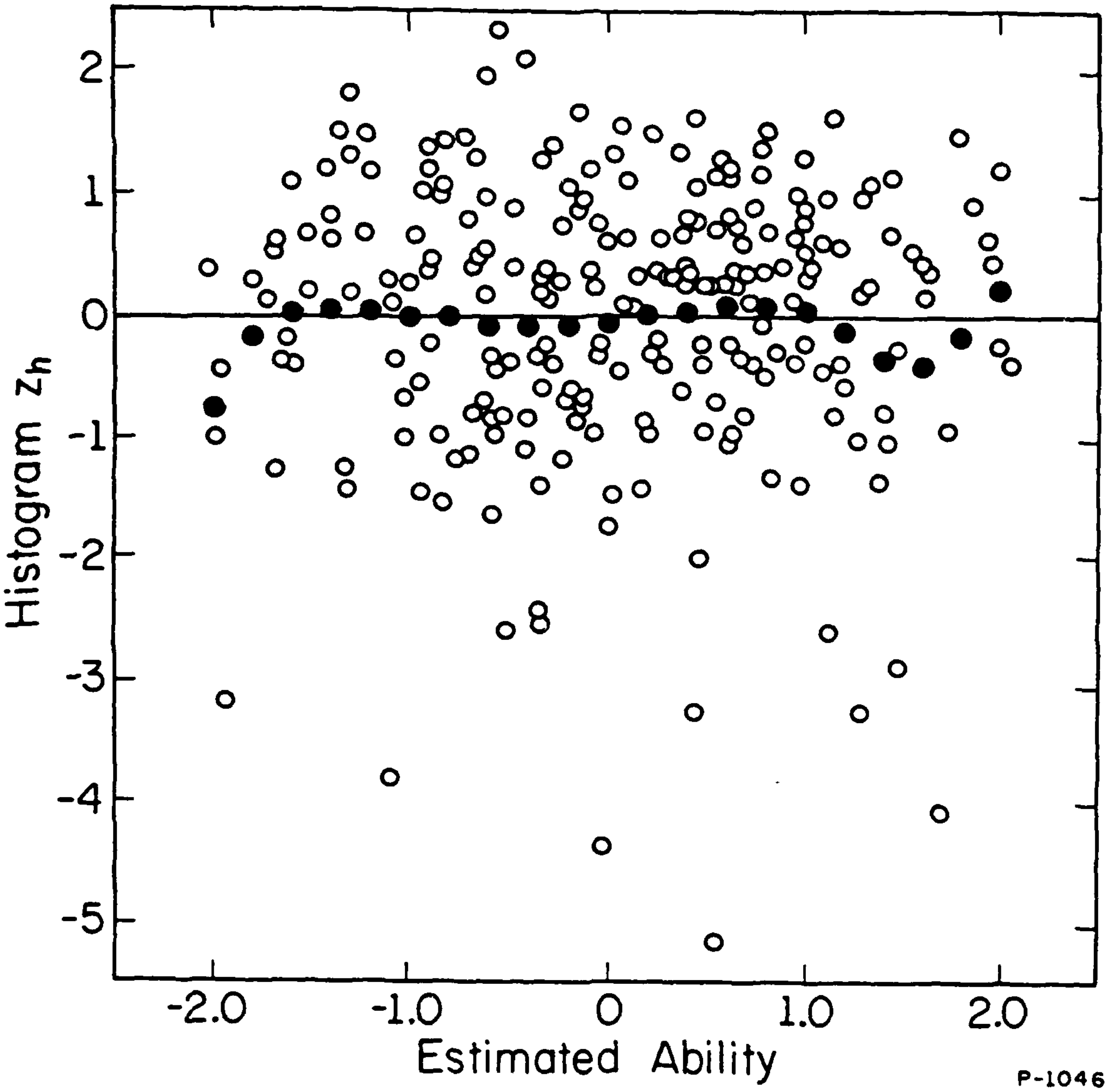
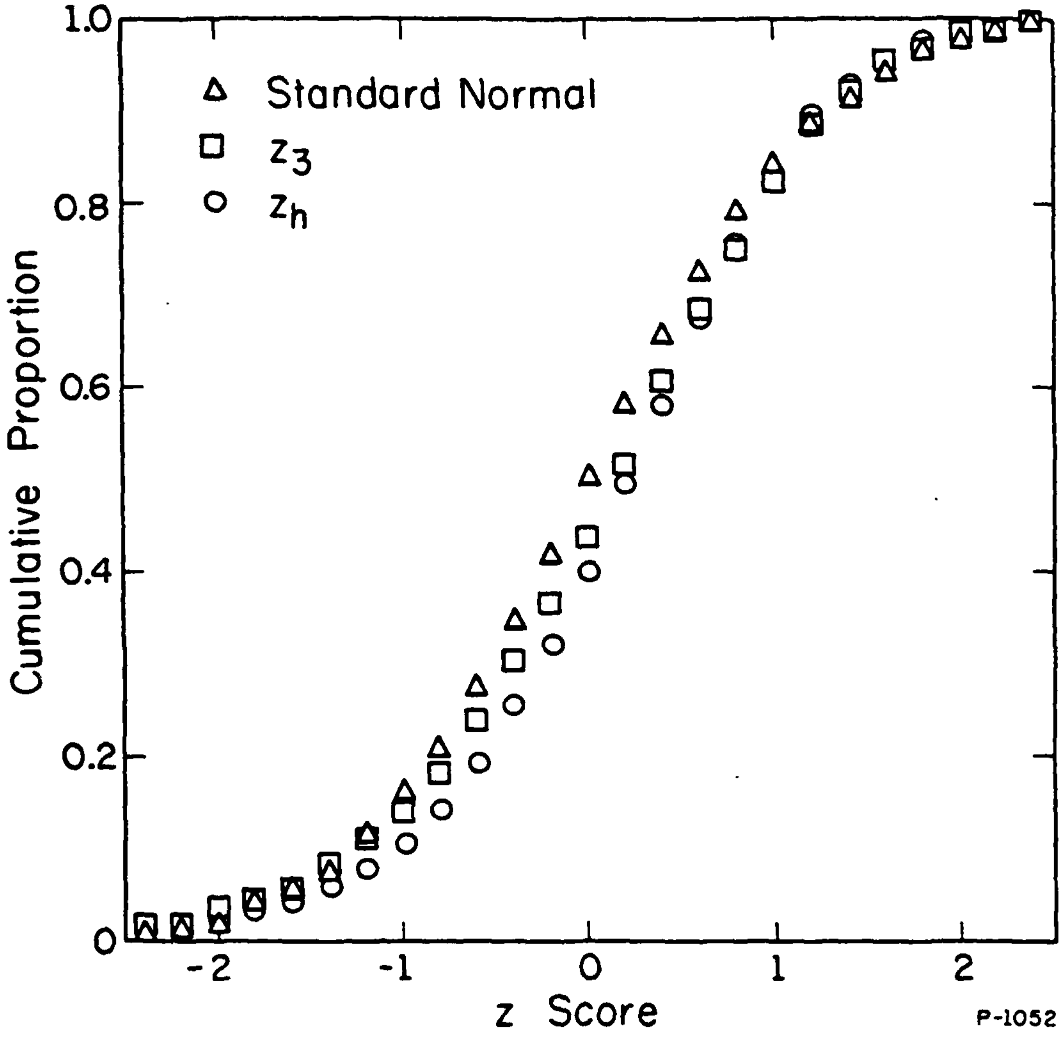

Three Parameter Logistic z_3
2
1
0
-1
-2
-3
-2.0
-1.0
0
1.0
2.0
Estimated Ability
P-1049

Fig.3



P-1046

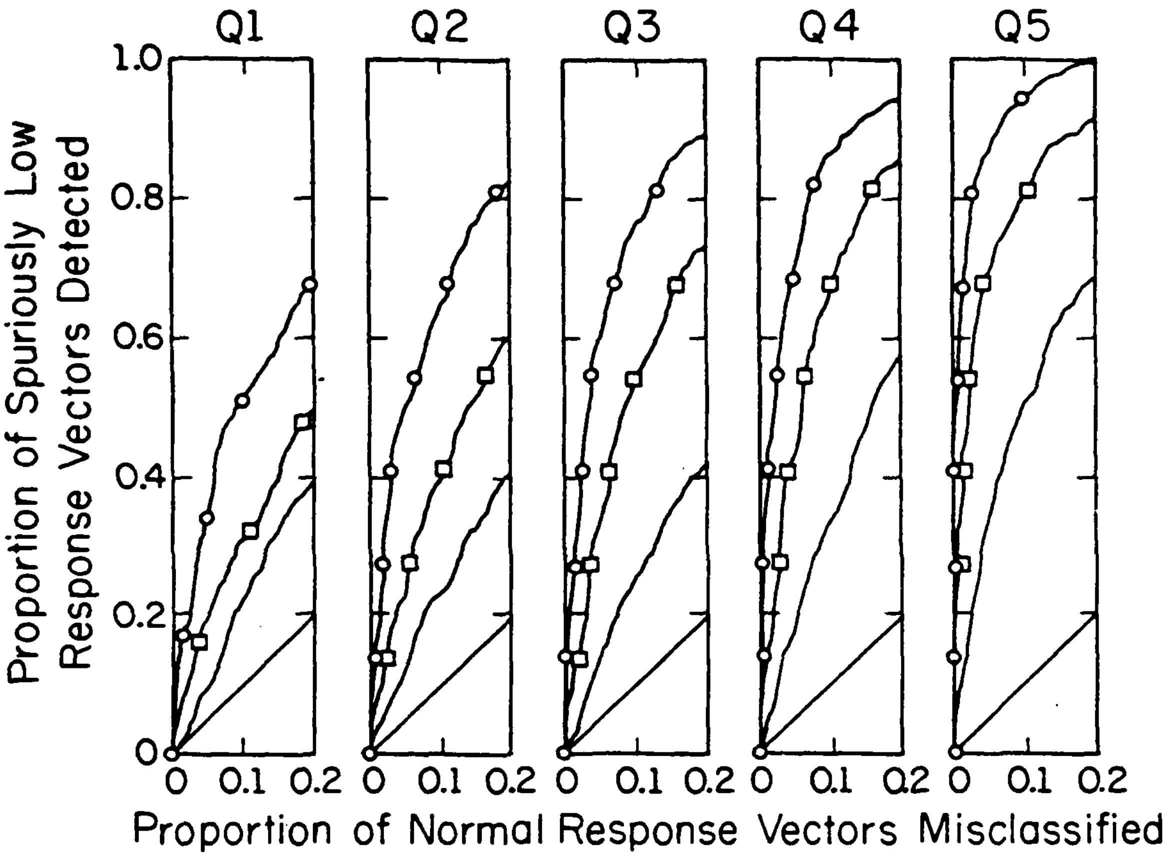
Fig. 4



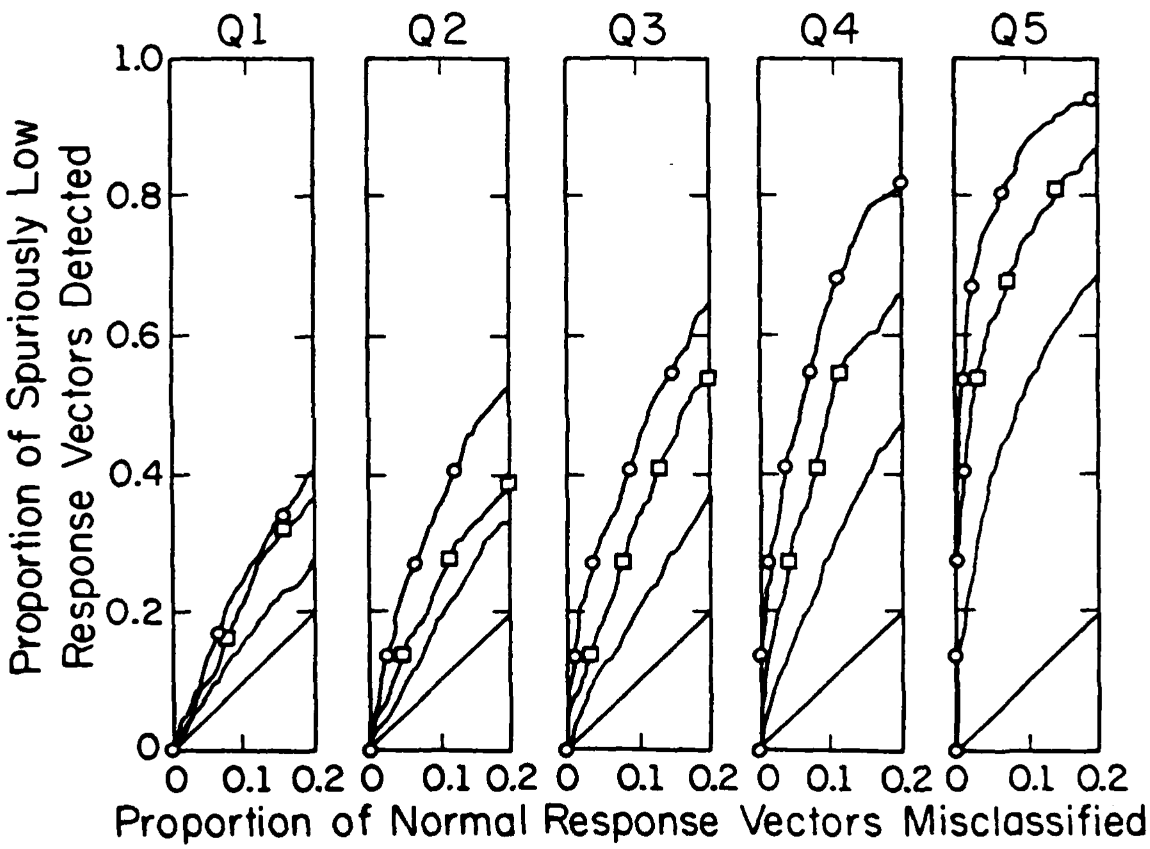
Standard Normal
z_3
z_h
Cumulative Proportion
z Score
1.0
0.8
0.6
0.4
0.2
0
-2
-1
0
1
2

P-1052

Fig. 5



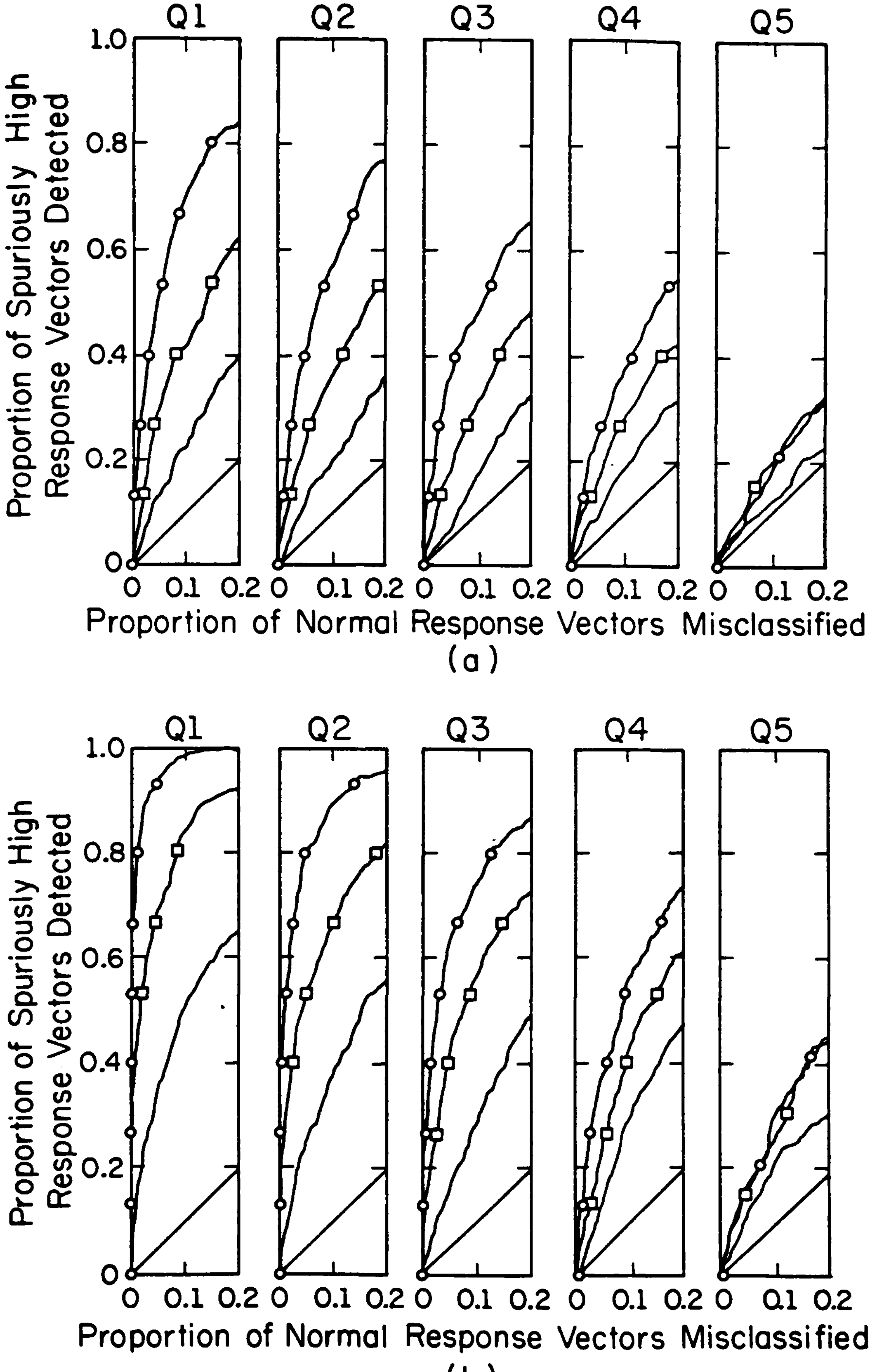
(a)



(b)

P-1050

Fig. 6



(a)

(b)

P-1051
Fig. 7

Navy

1 Dr. Ed Aiken
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Nick Bond
Office of Naval Research
Liaison Office, Far East
APO San Francisco, CA 96503

1 Lt. Alexander Bory
Applied Psychology
Measurement Division
NAMRL
NAS Pensacola, FL 32508

1 Dr. Robert Carroll
NAVOP 115
Washington , DC 20370

1 Dr. Stanley Collyer
Office of Naval Technology
800 N. Quincy Street
Arlington, VA 22217

1 CDR Mike Curran
Office of Naval Research
800 N. Quincy St.
Code 270
Arlington, VA 22217

1 Dr. John Ellis
Navy Personnel R&D Center
San Diego, CA 92252

1 DR. PAT FEDERICO
Code P13
NPRDC
San Diego, CA 92152

1 Dr. Cathy Fernandes
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

1 Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152

Navy

1 Dr. William L. Maloy (02)
Chief of Naval Education and Training
Naval Air Station
Pensacola, FL 32508

1 Dr. Kneale Marshall
Chairman, Operations Research Dept.
Naval Post Graduate School
Monterey, CA 93940

1 Dr. James McBride
Navy Personnel R&D Center
San Diego, CA 92152

1 Cdr Ralph McCumber
Director, Research & Analysis Division
Navy Recruiting Command
4015 Wilson Boulevard
Arlington, VA 22203

1 Dr. George Moeller
Director, Behavioral Sciences Dept.
Naval Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 06349

1 Dr William Montague
NPRDC Code 13
San Diego, CA 92152

1 Library, Code P201L
Navy Personnel R&D Center
San Diego, CA 92152

1 Technical Director
Navy Personnel R&D Center
San Diego, CA 92152

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

6 Personnel & Training Research Group
Code 442PT
Office of Naval Research
Arlington, VA 22217

1 LT Frank C. Petho, MSC, USN (Ph.D)
CNET (N-432)
NAS
Pensacola, FL 32508

Navy

1 Dr. Bernard Rimland (01C)
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Carl Ross
CNET-PDCD
Building 90
Great Lakes NTC, IL 60088

1 Mr. Drew Sands
NPRDC Code 62
San Diego, CA 92152

1 Dr. Robert G. Smith
Office of Chief of Naval Operations
OP-987H
Washington, DC 20350

1 Dr. Richard Snow
Liaison Scientist
Office of Naval Research
Branch Office, London
Box 39
FPO New York, NY 09510

1 Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Frederick Steinheiser
CNO - OP115
Navy Annex
Arlington, VA 20370

1 Mr. Brad Sympson
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. James Tweeddale
Technical Director
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Edward Wegman
Office of Naval Research (Code 411S&P)
800 North Quincy Street
Arlington, VA 22217

1 Dr. Douglas Wetzel
Code 12
Navy Personnel R&D Center
San Diego, CA 92152

Navy

1 DR. MARTIN F. WISKOFF
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 Mr John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152

Marine Corps

1 H. William Greenup
Education Advisor (E031)
Education Center, MCDEC
Quantico, VA 22134

1 Jerry Lehnus
CAT Project Office
HQ Marine Corps
Washington , DC 20380

1 Director, Office of Manpower Utilizatio
HQ, Marine Corps (MPU)
BCB, Bldg. 2009
Quantico, VA 22134

1 Headquarters, U. S. Marine Corps
Code MPI-20
Washington, DC 20380

1 Special Assistant for Marine
Corps Matters
Code 100M
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

1 DR. A.L. SLAFKOSKY
SCIENTIFIC ADVISOR (CODE RD-1)
HQ, U.S. MARINE CORPS
WASHINGTON, DC 20380

1 Major Frank Yohannan, USMC
Headquarters, Marine Corps
(Code MPI-20)
Washington, DC 20380

Army

1 Technical Director
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Kent Eaton
Army Research Institute
5001 Eisenhower Blvd.
Alexandria , VA 22333

1 Dr. Myron Fischl
U.S. Army Research Institute for the
Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Milton S. Katz
Training Technical Area
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Clessen Martin
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

1 Dr. William E. Nordbrock
FMC-ADCO Box 25
APO, NY 09710

1 Mr. Robert Ross
U.S. Army Research Institute for the
Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Robert Sasmor
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joyce Shields
Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Hilda Wing
Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333

Air Force

1 Technical Documents Center
Air Force Human Resources Laboratory
WPAFB, OH 45433

1 U.S. Air Force Office of Scientific Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, DC 20332

1 Air University Library
AUL/LSE 76/443
Maxwell AFB, AL 36112

1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

1 Mr. Raymond E. Christal
AFHRL/MOE
Brooks AFB, TX 78235

1 Dr. Alfred R. Fregly
AFOSR/NL
Bolling AFB, DC 20332

1 Dr. Patrick Kyllonen
AFHRL/MOE

Brooks AFB, TX 78235

1 Dr. Roger Pennell
Air Force Human Resources Laboratory
Lowry AFB, CO 80230

1 Dr. Malcolm Ree
AFHRL/MP
Brooks AFB, TX 78235

Department of Defense

12 Defense Technical Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC

1 Military Assistant for Training and Personnel Technology
Office of the Under Secretary of Defens
for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 Dr. W. Steve Sellman
Office of the Assistant Secretary
of Defense (MRA & L)
2B269 The Pentagon
Washington, DC 20301

1 Major Jack Thorpe
DARPA
1400 Wilson Blvd.
Arlington, VA 22209

1 Dr. Robert A. Wisher
OUSDRE (ELS)
The Pentagon, Room 3D129
Washington, DC 20301

Civilian Agencies

1 Dr. Vern W. Urry
Personnel R&D Center
Office of Personnel Management
1900 E Street NW
Washington, DC 20415

1 Mr. Thomas A. Warm
U. S. Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

1 Dr. Frank Withrow
U. S. Office of Education
400 Maryland Ave. SW
Washington, DC 20202

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

Private Sector

1 Dr. James Algina
University of Florida
Gainesville, FL 326

1 Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

1 1 Psychological Research Unit
NBH-3-44 Attn
Northbourne House
Turner ACT 2601
AUSTRALIA

1 Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND

1 Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

1 Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Tel Aviv, Ramat Aviv 69978
Israel

1 Dr. R. Darrell Bock
Department of Education
University of Chicago
Chicago, IL 60637

1 Dr. Robert Brennan
American College Testing Programs
P. O. Box 168
Iowa City, IA 52243

1 Dr. Glenn Bryan
6208 Poe Road
Bethesda, MD 20817

1 Bundministerium der Verteidigung
-Referat P II 4-
Psychological Service
Postfach 1328
D-5300 Bonn 1
F. R. of Germany

Private Sector

1 Dr. Ernest R. Cadotte
307 Stokely
University of Tennessee
Knoxville, TN 37916

1 Dr. John B. Carroll
409 Elliott Rd.
Chapel Hill, NC 27514

1 Dr. Norman Cliff
Dept. of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

1 Dr. Allan M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02138

1 Dr. Lynn A. Cooper
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 Dr. Hans Crombag
Education Research Center
University of Leyden
Boerhaavelaan 2
2334 EN Leyden
The NETHERLANDS

1 CTB/McGraw-Hill Library
2500 Garden Road
Monterey, CA 93940

1 Dr. Dattpradad Divgi
Syracuse University
Department of Psychology
Syracuse, NE 33210

1 Dr. Hei-Ki Dong
Ball Foundation
Room 314, Building B
800 Roosevelt Road
Glen Ellyn, IL 60137

1 Dr. Fritz Drasgow
Department of Psychology
University of Illinois
603 E. Daniel St.
Champaign, IL 61820

Private Sector

1 Dr. Susan Embertson
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
Lawrence, KS 66045

1 ERIC Facility-Acquisitions
4833 Rugby Avenue
Bethesda, MD 20014

1 Dr. Benjamin A. Fairbank, Jr.
McFann-Gray & Associates, Inc.
5825 Callaghan
Suite 225
San Antonio, TX 78228

1 Dr. Leonard Feldt
Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Dr. Richard L. Ferguson
The American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Univ. Prof. Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

1 Professor Donald Fitzgerald
University of New England
Armidale, New South Wales 2351
AUSTRALIA

1 Dr. Dexter Fletcher
University of Oregon
Department of Computer Science
Eugene, OR 97403

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 Dr. Janice Gifford
University of Massachusetts
School of Education
Amherst, MA 01002

Private Sector

1 Dr. Robert Glaser
Learning Research & Development Center
University of Pittsburgh
3939 O'Hara Street
PITTSBURGH, PA 15260

1 Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

1 DR. JAMES G. GREENO
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Ron Hambleton
School of Education
University of Massachusetts
Amherst, MA 01002

1 Dr. Delwyn Harnisch
University of Illinois
242b Education
Urbana, IL 61801

1 Dr. Paul Horst
677 G Street, #184
Chula Vista, CA 90010

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
603 East Daniel Street
Champaign, IL 61820

1 Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

1 Dr. Jack Hunter
2122 Coolidge St.
Lansing, MI 48906

Private Sector

1 Dr. Huynh Huynh
College of Education
University of South Carolina
Columbia, SC 29208

1 Dr. Douglas H. Jones
Advanced Statistical Technologies
Corporation
10 Trafalgar Court
Lawrenceville, NJ 08148

1 Professor John A. Keats
Department of Psychology
The University of Newcastle
N.S.W. 2308
AUSTRALIA

1 Dr. Scott Kelso
Haskins Laboratories, Inc
270 Crown Street
New Haven, CT 06510

1 CDR Robert S. Kennedy
Canyon Research Group
1040 Woodcock Road
Suite 227
Orlando, FL 32803

1 Dr. William Koch
University of Texas-Austin
Measurement and Evaluation Center
Austin, TX 78703

1 Dr. Stephen Kosslyn
1236 William James Hall
33 Kirkland St.
Cambridge, MA 02138

1 Dr. Alan Lesgold
Learning R&D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

1 Dr. Michael Levine
Department of Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Private Sector

1 Dr. Charles Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat 23
9712GC Groningen
Netherlands

1 Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

1 Mr. Phillip Livingston
Systems and Applied Sciences Corporatio
6811 Kenilworth Avenue
Riverdale, MD 20840

1 Dr. Robert Lockman
Center for Naval Analysis
200 North Beauregard St.
Alexandria, VA 22311

1 Dr. Frederic M. Lord
Educational Testing Service
Princeton, NJ 08541

1 Dr. James Lumsden
Department of Psychology
University of Western Australia
Nedlands W.A. 6009
AUSTRALIA

1 Dr. Don Lyon
P. O. Box 44
Higley , AZ 85236

1 Dr. Gary Marco
Stop 31-E
Educational Testing Service
Princeton, NJ 08451

1 Dr. Scott Maxwell
Department of Psychology
University of Notre Dame
Notre Dame, IN 46556

1 Dr. Samuel T. Mayo
Loyola University of Chicago
820 North Michigan Avenue
Chicago, IL 60611

Private Sector

1 Mr. Robert McKinley
American College Testing Programs
P.O. Box 168
Iowa City, IA 52243

1 Dr. Barbara Means
Human Resources Research Organization
300 North Washington
Alexandria, VA 22314

1
Professor Jason Millman
Department of Education
Stone Hall
Cornell University
Ithaca, NY 14853

1 Dr. Allen Munro
Behavioral Technology Laboratories
1845 Elena Ave., Fourth Floor
Redondo Beach, CA 90277

1 Dr. W. Alan Nicewander
University of Oklahoma
Department of Psychology
Oklahoma City, OK 73069

1 Dr. Donald A Norman
Cognitive Science, C-015
Univ. of California, San Diego
La Jolla, CA 92093

1 Dr. Melvin R. Novick
356 Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Dr. James Olson
WICAT, Inc.
1875 South State Street
Orem, UT 84057

1 Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

1 Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Cirle, NW
Washington, DC 20036

Private Sector

1 Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

1 Dr. James W. Pellegrino
University of California,
Santa Barbara
Dept. of Psychology
Santa Barabara , CA 93106

1 Dr. Mark D. Reckase
ACT
P. O. Box 168
Iowa City, IA 52243

1 Dr. Thomas Reynolds
University of Texas-Dallas
Marketing Department
P. O. Box 688
Richardson, TX 75080

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
Murray Hill, NJ 07974

1 Dr. Lawrence Rudner
403 Elm Avenue
Takoma Park, MD 20012

1 Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

1 Frank L. Schmidt
Department of Psychology
Bldg. GG
George Washington University
Washington, DC 20052

1 Dr. Walter Schneider
Psychology Department
603 E. Daniel
Champaign, IL 61820

Private Sector

1 Lowell Schoer
Psychological & Quantitative
Foundations
College of Education
University of Iowa
Iowa City, IA 52242

1 Dr. Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujusawa 251
JAPAN

1 Dr. Edwin Shirkey
Department of Psychology
University of Central Florida
Orlando, FL 32816

1 Dr. William Sims
Center for Naval Analysis
200 North Beauregard Street
Alexandria, VA 22311

1 Dr. Robert Sternberg
Dept. of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 Martha Stocking
Educational Testing Service
Princeton, NJ 08541

1 Dr. Peter Stoloff
Center for Naval Analysis
200 North Beauregard Street
Alexandria, VA 22311

1 David E. Stone, Ph.D.
Hazeltine Corporation
7680 Old Springhouse Road
McLean, VA 22102

1 Dr. William Stout
University of Illinois
Department of Mathematics
Urbana, IL 61801

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN
THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

Private Sector

1 Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

1 Dr. Kikumi Tatsuoka
Computer Based Education Research Lab
252 Engineering Research Laboratory
Urbana, IL 61801

1 Dr. Maurice Tatsuoka
220 Education Bldg
1310 S. Sixth St.
Champaign, IL 61820

1 Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

1 Dr. Douglas Towne
Univ. of So. California
Behavioral Technology Labs
1845 S. Elena Ave.
Redondo Beach, CA 90277

1 Dr. Robert Tsutakawa
Department of Statistics
University of Missouri
Columbia, MO 65201

1 Dr. V. R. R. Uppuluri
Union Carbide Corporation
Nuclear Division
P. O. Box Y
Oak Ridge, TN 37830

1 Dr. David Vale
Assessment Systems Corporation
2233 University Avenue
Suite 310
St. Paul, MN 55114

1 Dr. Kurt Van Lehn
Xerox PARC
3333 Coyote Hill Road
Palo Alto, CA 94304

1 Dr. Howard Wainer
Division of Psychological Studies
Educational Testing Service
Princeton, NJ 08540

Private Sector

1 Dr. Michael T. Waller
Department of Educational Psychology
University of Wisconsin--Milwaukee
Milwaukee, WI 53201

1 Dr. Brian Waters
HumRRO
300 North Washington
Alexandria, VA 22314

1 Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

1 Dr. Donald O. Weitzman
Mitre Corporation
1820 Dolley Madison Blvd
McLean, VA 22102

1 Dr. Christopher Wickens
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Dr. Rand R. Wilcox
University of Southern California
Department of Psychology
Los Angeles, CA 90007

1 German Military Representative
ATTN: Wolfgang Wildegrube
Streitkraefteamt
D-5300 Bonn 2
4000 Brandywine Street, NW
Washington , DC 20016

1 Dr. Bruce Williams
Department of Educational Psychology
University of Illinois
Urbana, IL 61801

1 Dr. Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

END

FILMED

DTIC